

Analisis Sentimen Pada Twitter Shopee Indonesia Menggunakan *K-Nearest Neighbor* Dengan Pembobot *Term Frequency-Inverse Document Frequency* (Tf-Idf)

Oleh: Fadhilah Az-zahra
Univeristas Muhammadiyah Semarang

Article history	Abstract
Submission :	Twitter is a social media that is in demand by many people and is transformed into a very fast information broadcaster today. The information released and circulating via the media is very free and has a lot of variety, such as news, opinions, questions, criticisms, comments, both positive and negative. Classification is a rule in text mining that collects content according to the similarity of the script. This rule allows tweets available on Twitter to be grouped together based on their type. For example, the substance of football, volleyball, and tennis is classified into the sports category. The classification procedure begins using preprocessing, then weighting the words, then categorization consisting of preprocessing and several stages, namely document cleaning, tokenizing, stopword removal, and stemming. The word weighting method used in this thesis is Term Frequency-Inverse Document Frequency (TF-IDF) and uses K-Nearest Neighbor (K-NN) with 80:20 training data. The KNN method is a classification of a set of data based on learning data that has been classified previously. Testing the accuracy of the classification of tweets on Twitter using the K-Nearest Neighbor (K-NN) method results in accuracy where the total data is 1300, with the k numbers entered being 1, 3, 5, and 7 and 9 respectively, the results are k = 1 , an accuracy of 77.31%; k = 3, accuracy 75.38%; k = 5, accuracy 75.38%; k = 7, 75.38% accuracy and k = 9 accuracy 75.38%.
Revised :	
Accepted :	
Keyword: <i>Preprocessing, K- Nearest Neighbor, Term frequency-inver document Frequency.</i>	

PENDAHULUAN

Perkembangan internet Indonesia kian hari semakin meningkat. Berdasarkan data survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) tahun 2016, pengguna Internet mencapai 132,7 juta orang dan mencapai 1.4326 juta pada tahun 2017 (news.detik.com). Hal ini juga terlihat dari survei APJII tentang penggunaan internet di berbagai bidang dimana pengguna internet cenderung membaca berita. Salah satunya seperti penggunaan Internet dalam bidang gaya hidup meraih hasil tertinggi dalam aktivitas media sosial mencapai 87,13%, kemudian unduhan musik 71,10%, unduhan film 70%, dan di ikuti bidang sosial politik. Hasil tertinggi yaitu kegiatan sosial membaca dan membaca. Lingkungan berkurang 50,26%. Terlihat dari survei ini bahwa penggunaan media sosial di

Indonesia memiliki tingkat penggunaan internet yang tinggi, diikuti oleh media baca massa online yang menggerakkan banyak Media massa mulai aktif di media sosial. (<https://apji.or.id/survei2017>).

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat. Minimumnya di dalam analisis sentimen terdapat 2 kategori dalam menentukan sentimen yaitu kategori sentimen positif dan negatif. Sentimen analisis sendiri merupakan salah satu model klasifikasi data dengan pendekatan supervised learning di dalam machine learning. Machine learning mempunyai 2 tipe teknik yaitu supervised learning dan unsupervised learning. (Brownlee, 2016)..

Supervised learning adalah salah satu tipe algoritma machine learning yang menggunakan data latih yang dikenal (training dataset) untuk membuat prediksi. Sedangkan unsupervised learning adalah salah satu tipe algoritma machine learning yang digunakan untuk menarik kesimpulan dari dataset yang terdiri dari input data labeled response. Metode unsupervised learning yang paling umum adalah analisa cluster, yang digunakan pada analisa data untuk mencari pola-pola tersembunyi atau pengelompokan dalam data, sedangkan pada supervised learning contoh yang paling umum adalah klasifikasi data. Mayoritas praktis dari machine learning sendiri menggunakan supervised learning. Karena besarnya pengaruh dan manfaat dari klasifikasi data menyebabkan penelitian dan aplikasi berbasis klasifikasi data seperti text mining, analisis sentiment justru berkembang pesat (Ghulam, 2016).

Salah satu penelitian terdahulu terkait K-Nearest Neighbor pernah dilakukan oleh Veverly Widyastuti (2014) untuk Sistem klasifikasi dokumen bahasa Jawa dengan Metode K-Nearest Neighbor (K-NN) “. Pada penelitian ini dilakukan pengujian dengan Cross Validation kemudian dilakukan uji presisi data yang digunakan sebanyak 40 dokumen dengan tingkat akurasi K=4 mencapai 95% , K=8 mencapai 92%.

Berdasarkan uraian di atas, peneliti ingin menerapkan metode K-Nearest Neighbor pada analisis sentiment di media Twitter dengan studi kasus Shopee Indonesia.

LANDASAN TEORI

Analisis Sentimen

Liu (2015) Analisis Sentimen merupakan Natural Language Processing yang bertujuan untuk membangun sebuah metode yang dapat diimplementasikan menjadi sebuah tools yang dapat digunakan untuk mengekstraksi informasi subjektif berupa sentimen atau opini dalam sebuah data text. kecenderungan penelitian tentang analisis sentimen berfokus pada pendapat yang menyatakan suatu sentimen memiliki nilai positif atau negatif.

Twitter

Adiyana dan Hakim (2015). Pada aplikasi twitter disediakan sebuah search engine yang dapat digunakan oleh pengguna twitter.

Dengan search engine tersebut kita bisa mendapatkan informasi terkait tweets yang ada pada aplikasi twitter. Informasi tersebut banyak digunakan untuk mencari wawasan terkait permasalahan yang sedang dihadapi. Informasi yang diberikan dapat berupa tweets, pengguna yang menuliskan tweets tersebut, lokasi tweets tersebut, dan lain-lain.

Shopee

miung, (2015). Saat ini, masyarakat Indonesia sudah banyak menggunakan teknologi informasi yang berbasis internet untuk memenuhi kebutuhannya. Perkembangan teknologi internet ini telah memberikan dampak positif bagi Indonesia dimana layanan internet sudah banyak dilakukan oleh individu, perusahaan, instansi pemerintahan maupun swasta Menggunakan internet bagi masyarakat Indonesia sudah menjadi hal yang menyatu dalam kehidupan sehari-hari. Terlebih lagi biaya untuk mengakses internet relatif murah, didukung pula dengan semakin murahnya harga ponsel pintar di Indonesia yang dijadikan sebagai penunjang kegiatan tersebut, hal tersebut, berdampak pula pada bertambahnya pengguna internet di Indonesia (Sumber: <https://www.emarketer.com/>,2015)..

Text Mining

Feldman dan Sanger (2007 dalam Fatimah 2018) Text Mining merupakan suatu proses penggalian informasi berdasarkan suatu sumber data dokumen yang berupa teks dalam suatu proses yang dilakukan dengan komputer. Text mining juga dikenal sebagai data mining text atau penemuan pengetahuan dari database tekstual. Menurut buku The Text Mining Handbook, text mining didefinisikan sebagai suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen data mining.

Klasifikasi

Suyanto(2017).Klasifikasi merupakan metode yang digunakan untuk mengelompokkan sebuah objek ke dalam kelompok atau kelas tertentu. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu Decision atau Classification Trees, Bayesian Classifiers atau Naïve Bayes classifiers, Neural Networks, Analisa Statistik, Algoritma Genetika, Rough Sets, K-Nearest Neighbor, Metode Rule Based, Memory Based

Reasoning, dan Support Vector Machines (SVM). Proses ini dilakukan agar data atau citra dapat dikategorikan dalam suatu kelas tertentu yang telah ditentukan

K-Nearest Neighbor

K- Nearest Neighbor (K-NN) adalah algoritma untuk mengklasifikasi objek baru berdasarkan atribut dan training samples (data latih). Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN. Algoritma K-NN menggunakan klasifikasi ketetangaan sebagai nilai prediksi dari sampel uji yang baru (Krisandi, dkk. 2013). Data latih akan dibangun dengan memperhatikan keseimbangan dokumen satu sama lain

$$\text{Cos}(i,k) = \frac{\sum k(di,dk)}{\sqrt{\sum k d^2 ik} \sqrt{\sum k d^2 jk}}$$

Keterangan:

$\sum k(di, dk)$ = vector dot produk dari i ke k

$\sum k d^2 jk$ = panjang vektor i

Algoritma K-NN (Krisandi, dkk. 2013) adalah algoritma yang menentukan nilai jarak pada pengujian data *testing* dengan data *training* berdasarkan nilai terkecil dari nilai ketetangaan terdekat

$$D_{nn}(c1, c2) = \min_{1 \leq I \leq r, 1 \leq j \leq s} d(y_i, z_j)$$

Confusion Matrix

Dalam mengukur tingkat akurasi sistem orientasi sentimen, maka digunakan tabel confusion matrix. Confusion matrix adalah sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan.

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN+NF+FN}$$

Keterangan :

TP = jumlah dokumen dari kelas 1 yang benar dan diklasifikasikan sebagai kelas 1.

TN = jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0.

FP = jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.

FN = jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

NP = jumlah dokumen dari kelas 0 yang bersifat netral

NF = jumlah dokumen dari kelas 0 yang bersifat Netral

METODE PENELITIAN

Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data komentar dari Shopee Indonesia yang diambil melalui Twitter selama event shopee berlangsung. data yang didapatkan dari hasil Crawling menggunakan Data dari API pada Twitter. dengan melakukan hasil klasifikasi K- Nearest Neighbor dengan menggunakan pembobot TF-IDF serta model data latih seimbang sentimen positif, negatif, dan netral, serta menghitung hasil akurasi. kemudian melakukan observasi data dari Twitter API tentang komentar netizen terhadap E-Commerce Shopee Indonesia yang diambil pada media Twitter. pengambilan data memanfaatkan fitur dari twitter yang diperuntukkan untuk developer yang terdapat pada website <https://developers.twitter.com/>. Setelah data didapatkan dari twitter API dalam bentuk XML, kemudian data tersebut disimpan langsung dengan menggunakan format CSV.

Variabel dan Struktur Data

Tabel 1 Variabel Penelitian

No	Variabel	Keterangan
1	$Y_{1,t}$	Uang kartal pada bulan ke-t
2	$Y_{2,t}$	Uang giral pada bulan ke-t

Data yang didapatkan dari hasil Crawling data menggunakan API twitter dibagi menjadi data training dan data testing menggunakan perbandingan 80% : 20%. Untuk struktur data yang digunakan dalam penelitian ini setelah dilakukan preprocessing pada data teks tweet dari variabel prediktor yaitu kata dasar setiap tweet dan variabel respon yaitu klasifikasi sentiment tweet (positif dan negatif).

Langkah Penelitian

Langkah-langkah dalam penelitian ini yaitu :

1. Mengambil data tweet dengan menggunakan Twitter API
2. Memasukkan keyword Shopee44
3. Menyimpan hasil Crawling ke database
4. Text Preprocessing
5. Lalu dilakukan seleksi dengan TF-IDF yaitu mengubah teks menjadi vektor dan pembobotan kata
6. Ambil data dari hasil Crawling

HASIL PENELITIAN dan PEMBAHASAN

Crawling Data Twitter

Crawling data merupakan tahap awal berupa pengumpulan data untuk diproses menuju tahap text mining. Dalam proses crawling data penggunaan twitter API sebagai sarana dan pintu untuk mengambil data berupa tweet dari media sosial twitter.

Crawling Data Sebagai berikut:

1. Koneksi API merupakan tahap awal untuk melakukan crawling data. Koneksi API digunakan oleh aplikasi dengan cara menghubungkan database dengan twitter API.
2. Registrasi adalah tahap kedua dari conceptual model crawling data. Registrasi yang dimaksud setelah tahap koneksi API berupa registrasi akun sebagai developer twitter untuk memperoleh akses penggunaan twitter API. Website untuk registrasi adalah developer.twitter.com.
3. Otentikasi adalah tahap ketiga dari conceptual model crawling data. Setelah berhasil melakukan registrasi sebagai developer twitter pada website developer.twitter.com, pengguna dapat membuat access kepada twitter API dengan menggunakan 4 kode yang diberikan untuk 1 aplikasi
4. API key, API secret key, Access token, Access token secret. Dengan menggunakan akses web service yang diberikan pengguna dapat mengambil sumber data baik tweet, tanggal, bahkan nama akun dengan bahasa

Hasil Preprocessing

Pada tahapan text preprocessing digunakan untuk melakukan seleksi data untuk mengubah data teks yang tidak terstruktur atau sembarang menjadi data yang terstruktur

Tabel 3. Hasil Case Folding

Sebelum	Sesudah
@shopee44 skrg late night Sale puncak Shopee mega Sale	@shopee44 skrg late sale puncak shopee mega sale

Dari Tabel di atas diperoleh Merupakan tahap awal setelah teks (komentar pada twitter) berupa hasil crawling didapatkan. Teks diproses oleh aplikasi dengan merubah karakter menjadi huruf kecil.

Tabel 4. Hasil *Filtering*

Sebelum	Sesudah
@shopee44 skrg late night Sale puncak Shopee mega Sale	Sekarang late sale puncak shopee mega sale

Berdasarkan Tabel 4, Merupakan tahap penghapusan URL ,Hastag pada dokumen. Berikut gambaran hasil Filtering.

Tabel 5. Hasil *Tokenizing*

Sebelum	Sesudah
@shopee44 skrg late night Sale puncak Shopee mega Sale	Sekarang,late,night,sale, puncak, shopee,mega,sale

Berdasarkan Tabel 5, Merupakan tahap dimana dalam satu dokumen dipecah menjadi token dengan dibatasi spasi. Berikut hasil *Tokenizing*.

Normalisasi Data

Normalisasi dilakukan dua kali. Pertama normalisasi kata yang memiliki kata berlebih dan tidak sesuai dengan standar ejaan KBBI. Kedua normalisasi kata tidak baku yang diubah menjadi kata baku sesuai dengan ejaan KBBI

Stopword

Merupakan proses penghapusan kata yang tidak penting dalam text. Hal tersebut dilakukan untuk meningkatkan akurasi dari pembobotan term dan memperkecil jumlah kata disetiap dokumen. Pada tahap ini kata yang ada dalam kamus stopwords akan dibuang hal ini dikarenakan stopwords dianggap sebagai kata tidak penting atau tidak berguna dan tahapan ini juga dapat menekan penggunaan waktu dalam proses pembobotan kata. (Tala,2003)

Stopword didapatkan dari <https://github.com/stopwords-iso/stopwords-id>

Tabel 6. Hasil Stemming

Sebelum	Sesudah
di Shopee 44 grtis ongkir tanpa batas belanja	shopee 44 gratis ongkir tanpa batas belanja

Berdasarkan Tabel 6, Jika semua langkah telah dilakukan tetapi kata dasar tersebut tidak ditemukan pada kamus, maka algoritma ini mengembalikan kata yang asli sebelum dilakukan tahapan Stemming.

Pelabelan Data Sentimen

Setelah melalui proses preprocessing, selanjutnya akan dilakukan analisis sentimen untuk pelabelan data. Pembobotan kata dilakukan dengan menghitung frekuensi kemunculan kata pada sebuah dokumen teks. Semakin sering sebuah kata muncul pada sebuah dokumen teks, maka bobot kata tersebut semakin besar dan kata tersebut dianggap sebagai kata yang sangat merepresentasikan dokumen teks tersebut (Yates dan Neto, 1999 dikutip dalam Basnur, 2009).

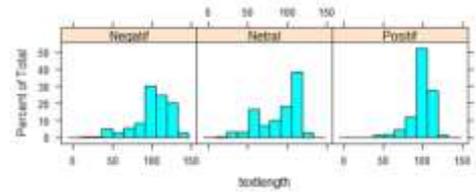
Pada pelabelan kelas sentimen menggunakan kategori positif dan kategori negatif, dengan jumlah data latih 80:20 adapun hasil dari pelabelan untuk Shopee44 dapat dilihat sebagai berikut

Jumlah Sentimen Positif dan Negatif



Gambar 1. Pelabelan sentiment positif dan Negatif

Berdasarkan gambar 1. didapatkan bahwa jumlah sentiment positif sebesar 359 tweets dan sentiment negatif sebesar 345 tweets.



Gambar 2 pelabelan Sentimen

Berdasarkan hasil pelabelan kelas sentimen didapatkan bahwa sentiment positif untuk Shopee44 sebanyak 359 tweets sedangkan untuk sentiment negatif Shopee44 sebanyak 345 tweets. Tweets diklasifikasikan sebagai sentiment positif apabila mengandung pernyataan positif seperti pujian, ungkapan terima kasih, atau testimoni positif tentang Shopee44. Suatu ulasan diklasifikasikan sebagai sentiment negatif apabila mengandung kalimat negatif seperti ketidakpuasan, ketidaknyamanan, penghinaan, laporan, layanan.

Tabel 6. Pelabelan Sentimen

Tabel 6. Simulasi Perhitungan Skor Sentimen

Teks Ulasan	Kata Positif	Kata Negatif
Shopee puncak pengiriman cepat tunda suka chat respon belanja baik produk pesanan sesuai pilihan harga murah pengiriman cepat	Cepat suka baik Sesuai Murah Cepat	tunda
Jumlah	6	1

Berdasarkan Tabel 6 diperoleh perhitungan sebagai berikut :

Skor = (Jumlah kata positif) – (Jumlah kata negatif) Skor

Skor = 6-1

Skor = 5

Skor akhir yang diperoleh dari simulasi perhitungan bernilai > 0, sehingga hasil ulasan bersifat Positif.

Pembobotan dengan Term Frequency-Inverse Document Frequency (TF-IDF)

Untuk simulasi TF-IDF akan dijelaskan perhitungan bobot dokumen terhadap query sesuai dengan rumus yang telah dijelaskan. Contoh query terms (Q) yang digunakan adalah “shopeealdebaranandin”, “Shopeepuncak”, “sale”. Misal terdapat tiga buah dokumen yaitu: Dokumen 1 (d1) = shopeealdebaranandin.

Dokumen 2 (d2) = ShopeePuncak. Dokumen 3 (d3) = sale.

Tabel 7. Contoh Query Terms (Q)

Text	Text	text
Shopeealdebaranandin	Shopeepuncak	gratis
Shopeealdebaranandin	Baru	Sale
ikatan cinta	Ongkir	Bisa

Keterangan :

Dokumen 1 (d1) = shopeealdebaranandin

Dokumen 2 (d2) = ShopeePuncak

Dokumen 3 (d3) = Sale

Adapun beberapa term (document term) yang didapatkan dari ketiga dokumen tersebut yaitu :

a. Jumlah kemunculan kata "murah" dalam d2 adalah sebanyak satu kali sehingga $tf_j = tfsale = 1$

b. Total seluruh dokumen yang ada yaitu sebanyak tiga dokumen sehingga $D = 3$

c. Dari ketiga dokumen tersebut, hanya satu dokumen yang memuat kata "murah" sehingga $df_j = dfsale = 1$

$$[Wi]_{-j} = ([tf]_{-j} \times \log) D/df_j$$

$$(Sale) = 1 \times \log \frac{3}{1}$$

$$(Sale) = 0,477$$

Klasifikasi

Tabel 8 Accuracy

Nilai K	Accuracy
K=1	77,31 %
K=3	75,38%
K=5	75,38%
K=7	75,38%
K=9	75,38%

Pada tahap ini dilakukan evaluasi pada klasifikasi menggunakan metode K-Nearest Neighbor. Pada table 8 diatas dapat diketahui nilai accuracy yang didapatkan dari masing-masing nilai K dengan data testing 80:20 menggunakan metode K-Nearest Neighbor didapatkan hasil dari K=1 sebesar 77,31% K=3 sebesar 75,38% K=5 sebesar 75,38% K=7 sebesar 75,38% dan K=9 sebesar 75,38%. Dari nilai accuracy yang didapatkan dari masing-masing kelas atau ketertangan didapatkan nilai accuracy terbesar yaitu K=1 sebesar

77,31%. Maka dapat dikatakan bahwa nilai accuracy terbaik pada masing-masing K atau ketertangan terdekat dari K=1 K= 3 K=5 K=7 K=9 adalah K=1 yaitu sebesar 77,31%

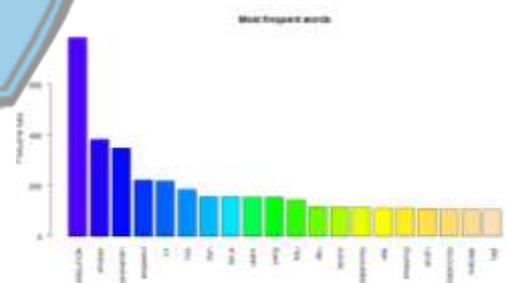
Tabel 9. Confusion Matrix Klasifikasi k=1

Prediksi	negatif	Netral	Positif
Negatif	38	7	0
Netral	26	124	24
Positif	0	2	39

Sentimen Ulasan Positif

Data ulasan positif yang digunakan adalah data hasil pelabelan yang dilakukan baik menggunakan kamus lexicon maupun secara manual. Ekstraksi informasi pada ulasan positif dilakukan secara berulang-ulang hingga mendapatkan informasi tentang ulasan positif pengguna Bukalapak yang paling sering diulas / dibicarakan. Ulasan positif tersebut diidentifikasi berdasarkan frekuensi kata dalam ulasan, berikut adalah visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan pengunjung dengan klasifikasi ulasan positif.

Input model VAR-SVR yang digunakan adalah variabel yang signifikan pada model terbaik VAR yaitu VARIMA (3,1,0). Sehingga lag-lag yang signifikan pada model tersebutlah yang dijadikan input pemodelan VAR-SVR. Pemodelan VAR-SVR dilakukan secara terpisah pada masing-masing data.



Gambar 2 sentimen ulasan positif

Pada hasil klasifikasi ulasan positif, dari jumlah ulasan positif sebanyak 1300 ulasan, diperoleh beberapa kata yang paling banyak muncul diantaranya adalah kata "shopeepuncak" dengan frekuensi sebanyak 650 kali, "shopee" sebanyak 400 kali dan seterusnya. Kata-kata yang muncul pada Gambar 4.5 merupakan kata yang memiliki sentiment positif dan merupakan topik pembicaraan yang paling banyak diulas oleh pengguna tweets pad event Shope44

SIMPULAN dan SARAN

Simpulan

Berdasarkan hasil analisis yang sudah dijelaskan, didapatkan beberapa kesimpulan sebagai berikut:

1. Hasil pengklasifikasian sentimen Shopee44 dengan data latih 80:20 menggunakan metode K-Nearest Neighbor didapatkan jumlah sentimen positif sebanyak 359, sentiment negative sebanyak 345 dan netral sebanyak 596.
2. Didapatkan hasil akurasi K=1 sebesar 77.31% K=3 sebesar 75.38% K=5 sebesar 75.38% K=7 sebesar 75.38% dan K=9 sebesar 75.38%. Berdasarkan hasil pengklasifikasian sentimen menggunakan K-Nearest Neighbor didapatkan nilai akurasi terbaik yaitu pada K=1 sebesar 77.31%

Saran

1. Penelitian ini hanya menganalisis satu e-commerce, untuk penelitian selanjutnya dapat membandingkan lebih dari satu e-commerce atau menganalisis model e-commerce lain.
2. Data yang digunakan pada penelitian ini hanya satu periode pada system tweets Shopee44, sehingga perlu ditambahkan data agar hasil klasifikasi sentimen lebih baik.

Daftar Pustaka

Agnes, K. (2010). Analisis Dan Perancangan Business Intelligence Untuk Seat Load Factor Pada Pt. Garuda Indonesia. DKI Jakarta: Universitas Bina Nusantara.

Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. Jurnal Teknologi Indonesia .

Badu, Z. S. (2016). Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Dana Desa. Jurnal Informatika, Gorontalo, 8.

DavidHaryalesmana.2021.<https://github.com/masdevid/IDOpinionWords/blob/master/positive.txt> Diakses pada 4 April 2021.

Dedy Santoso, Dian Eka Ratnawati, dan Indriati (2015) Perbandingan Kinerja Metode Naïve Bayes, K-Nearest Neighbor, dan Metode

Gabungan KMeans dan LVQ dalam Pengkategorian Buku Komputer Bahasa Indonesia Berdasarkan Judul dan Sinopsis

Feldman, R., & Sanger, J. (2007). *Advanced Approaches in Analyzing Unstructured Data, The Text Mining Handbook*. New York: Cambirdge UP.

Hotho A. (2005). *A Brief Survey of Text Mining*. University of Kassel.

Brownlee, Jason 2016. *Supervised Learning and Unsupervised Learning*. <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>. Diakses pada 10 April 2018.

Feldman, R., & Sanger, J. (2007). *Advanced Approaches in Analyzing Unstructured Data, The Text Mining Handbook*. New York: Cambirdge UP.

Hotho A. (2005). *A Brief Survey of Text Mining*. University of Kassel.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. (H. Graeme, Ed.) (1st ed.). Chicago : Morgan & Claypool Publisher. Retrieved from <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>

Liu, B. (2015). *Sentiment Analysis: Opinion Mining, Sentiment, and Emotions*. (H. Graeme, Ed.) (1st ed.). Chicago : Morgan & Claypool Publisher. Retrieved from <https://www.cs.uic.edu/~liub/FBS/chapter-1-and-chapter-2.pdf>

Matulatuwa, F. M., Studi, P., Sistem, M., Informasi, F. T., Kristen, U., & Wacana, S. (2017). *Text Mining dengan Metode Lexicon Based untuk Sentiment Analysis Pelayanan PT. POS Indonesia Melalui Media Sosial Twitter Jurnal*, (September).

Oktinas, Willa (2017). Analisis Sentimen Pada Acara Televisi Menggunakan Improved K-Neasrest Neighbor Tugas Akhir Teknik Informatika. Medan: Universitas Sumatra UtaraMigas Di Indonesia.*Prosiding Statistika Vol 6, No. 2 Tahun 2020*. 41-48

Veverly Widyastuti (2017). Sistem klasifikasi dokumen bahasa Jawa dengan Metode K-

Nearest Neighbor (K-NN) Tugas Akhir
Teknik Informatika. Medan: Universitas
Sumatra Utara



